IMPROVED AIR POLLUTION FORECASTING WITH HYBRID MACHINE LEARNING ALGORITHMS

By

Dhanalakshmi M

(Reg. No: 19PHCSF007)

Supervisor

Dr. V. Radha

Professor

Synopsis submitted to

Avinashilingam Institute for Home Science and Higher Education for Women

Coimbatore – 641043

In partial fulfilment of the requirements for the Degree of

Doctor of Philosophy in Computer Science

SEPTEMBER 2023

Improved Air Pollution Forecasting with Hybrid Machine Learning Algorithms

1. Introduction

Air pollution is the major causes of global warming which leads to more destructive natural calamities such as cyclones and storms. If we can able to predict the air pollution, there can be the solution to stop the cause of air pollution for avoiding more dangerous situations caused by air pollution. So, several machine learning methods are developed for forecasting air pollution with quality prediction. Data-driven prediction models usually achieve better forecasting of air pollution and fail to implement the prediction modeling successfully. To implement the model successfully for predicting the air pollution with huge volume of data and with multiple number of features, the enhanced machine learning techniques has to be identified. To that extent, this research work is developed with three enhanced models which contain processes like preprocessing, feature selection and classification.

Deep-AIR which uses the CNN and LSTM construction was performed in (Qi Zhang *et al.,* 2022) to fill the gap for achieving the fine-grained city-wide air pollution evaluation using domain-specific features to capture the Spatio-temporal features. Variational Auto Encoder based on the innovative Integrated Multiple Directed Attention Deep Learning architecture [VAE-IMDA] was introduced in (Abdelkadar Dairi *et al*., 2021) to development of conventional VAE and attention method to predict diverse air pollutants in a computationally efficient and accurate way. Also, temporal dependencies among non-linear approximations focused on the relevant feature extraction were attained. In addition, the authentication was used for univariate and multivariate with the help of pollution time series data prediction.

To improve the prediction of air pollution by using Convolutional neural network-Long Short-Term Memory, Sparse Denoising Autoencoder (CNN–LSTM–SDAE) (CLS) model was presented in Krishna Rani Samal *et al.,* 2021). But it was unsuccessful in enhancing the multivariate performance of prediction. A hybrid intelligent technique based on LSTM and multi-verse optimization algorithm (MVO) was proposed in (AzimHeydari *et al*., 2022) for forecasting and detecting air pollution. However, the accurate prediction was not developed.

An advanced deep convolutional network was carried out in (Gao Huang *et al*., 2021) to the huge-scale prediction of air pollution. However, the designed method was unsuccessful in measuring the various factors to increase the prediction accuracy. A deep learning solution termed CNN-LSTM was performed in (AbdellatifBekker *et al*., 2021) to predict the PM2.5 concentration. But the designed method was unsuccessful in utilizing the long-term *p*rediction model to increase the accuracy.

To carry out the accurate prediction of air pollution, using CWFM was proposed in (Bingchun Liu *et al*., 2021). But the evaluation of accurate pollution was not developed to the time and individual capacity constraints. Deep-learning construction was achieved in (Philipp

Hahnel *et al.,* 2020) to the air-pollution monitoring and predicting the capability and reduce the run-time. However, the complexity of time and space consumption was not detected.

The predictive method was carried out in (Doreswamy *et al*., 2020) to predict the awareness of atmospheric air and reduce the mean square error. However, the data preprocessing of the dataset was not achieved to reduce the time consumption. The intelligent hybrid air-quality prediction model was intended in (JianZhou Wang *et al*., 2021) to make accurate air-quality prediction data with climate-influencing features. However, it failed to improve the prediction of air quality.

The 1D-CNNs and Bi-LSTM were performed in (Shengdong Du *et al.,* 2021) to take out the spatial correlation features of prediction air quality. But multi-step prediction capability and enhanced outcomes were not achieved in several prediction circumstances. The novel deep learning framework was performed in (Ning Jin *et al.,* 2021) to combine MTMC-NLSTM for accurate forecasting of air quality. But the time complexity investigation was not focused.

A classified deep learning scheme was proposed in (S Abirami and P Chitra, 2020) for the prediction of air quality. But ensemble methods were not used to enhancement in prediction performance. MMSL were implemented in (DewenSeng *et al*.,2021) to forecast air quality. Yet, the error data during the forecasting was not reduced.

The spatial-attention embedded recurrent neural network was implemented in (Yu Huang *et al.*,2021) for the forecasting of air quality prediction using dynamic Spatio-temporal associations. But it was unsuccessful in reducing the error and the complexity of the algorithm was not minimized. The LSTM air quality prediction method was developed in (XiangyuZou *et al*.,2021) to establish the spatiotemporal attention technique. However, the quality forecasting was not enhanced.

An optimizing BP neural network established improved particle swarm optimization was used in (Yuan Huang *et al.,* 2020) to attain better accuracy. But the difficulty analysis was not focused on. A method based on integrated dual LSTM developed in (HongqiaoChen *et al.,* 2021) to reduce the hardness of forecasting. But it was hard to develop the accuracy of different data predictions.

DAEDN scheme was proposed in (JianXianCai *et al.,* 2020) to reduce noise and improve the air quality forecasting accuracy. However, it was not effective in reducing the space complexity due to the processing of huge amounts of data. The multi-scale construction was performed in (Shan Jiang *et al*., 2021) for term wise prediction of air quality. However, the designed method was unsuccessful in maintaining the relevant factors for the enhancement of air quality forecasting. The Research Problem is thus formulated as below,

"Design and implement a hybrid machine learning-based forecasting model that leverages multi-modal data to provide real-time predictions of various air pollutants with spatial and temporal features."

In order to find the solution to the above problem statement, the primary objectives of the research work are to improve the performance of air pollution forecasting accuracy, time and error rate in terms of multiple number of features with time and space complexity. The specific objectives framed are,

- To Develop Linear Regression and Multiclass Support Vector (LR-MSV) Model
- To Develop Bilateral Transformative Broken-Stick Regression-based Quadratic Weighted Emphasis Boost Classification (BTBSR-QWEBC) Model
- To Develop Discretized Regression and Least Square Support Vector (DR-LSSV) Model

2. Research Methodology

The research methodology of three proposed models is developed for improving forecasting accuracy on air pollution with minimum error rate and time. Figure 1 illustrates the methodology of three different proposed models.

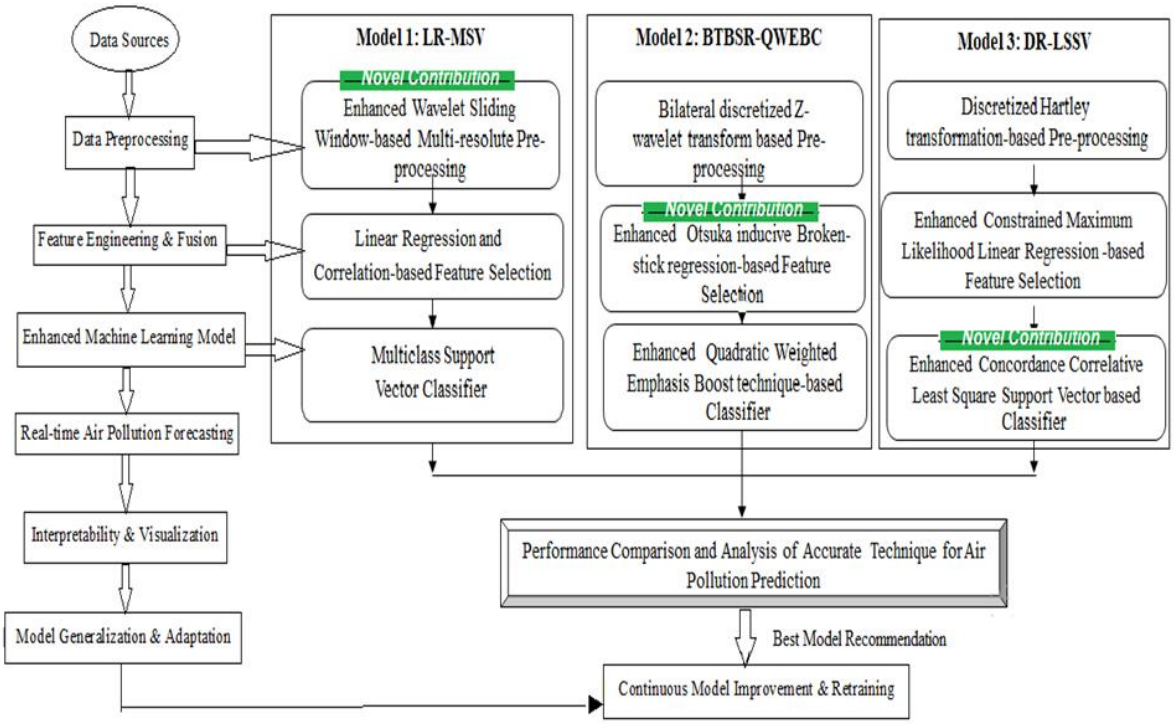
Data Sources
Data Preprocessing
Feature Engineering & Fusion
Enhanced Machine Learning Model
Real-time Air Pollution Forecasting
Interpretability & Visualization
Model Generalization & Adaptation
Model 1: LR-MSV
Novel Contribution
Enhanced Wavelet Sliding Window-based Multi-resolute Pre-processing
Linear Regression and Correlation-based Feature Selection
Multiclass Support Vector Classifier
Model 2: BTBSR-QWEBC
Bilateral discretized Z-wavelet transform based Pre-processing
Novel Contribution
Enhanced Otsuka inducive Broken-stick regression-based Feature Selection
Enhanced Quadratic Weighted Emphasis Boost technique-based Classifier
Model 3: DR-LSSV
Discretized Hartley transformation-based Pre-processing
Enhanced Constrained Maximum Likelihood Linear Regression -based Feature Selection
Novel Contribution
Enhanced Concordance Correlative Least Square Support Vector based Classifier
Performance Comparison and Analysis of Accurate Technique for Air Pollution Prediction
Best Model Recommendation
Continuous Model Improvement & Retraining

Figure 1: Overall Proposed Methodology

The designed models are LR-MSV, BTBSR-QWEBC, and DR-LSSV. These models are presented to forecast with a lesser error rate. Initially, the LR-MSV model is proposed to predict the air quality for enhanced forecasting accuracy. At a particular time, entire air quality data is failed to predict. It consumed large memory space for storing air data. To forecast air data with minimum memory consumption, the BTBSR-QWEBC model is designed in the next phase. The regression similarity coefficient is difficult to extract relevant features with a minimum error rate. At last, the DR-LSSV model is improved for achieving the efficient and robust features of data to attain better forecasting accuracy of air pollution

with minimum time and minimum error rate. A detailed explanation of three proposed models is presented in the subsection below.

2.1 LR-MSV Model

The LR-MSV Air Pollution Forecasting Model is designed to not only anticipate air quality data, but also to use AQI measurements to lead the path for air pollution control. The respiratory system is just one of the human body parts that are impacted by air pollution (AP), which also has an impact on the environment. In order to assess the AP level and predict the pollutant concentrations correctly while utilizing less noise, systematic Air Quality (AQ) forecasting is required. The wavelet decomposition method is a widely used signal processing approach that is specifically used in time series prediction. It is crucial to presumptively use a combination of progressions, superior frequency feature methods, and low-frequency features to decompose the non-smooth discrete time series air quality information. The frequency of improved feature mechanisms depends on the wavelet decomposition's number of layers. In this study, the raw air quality sensor data is processed using a Wavelet Sliding Window-based Multi-resolute preprocessing model that takes both time and frequency-based characteristics into account. In order to achieve the preprocessing through the sliding window to the noise removal, multi-resolute is defined as the time and frequency. Every time an instance occurs, the sliding window is established, and the samples use the values accordingly.

The next step is to select the best features from the preprocessed air quality data. Based on the correlation between the characteristics, linear regression and correlation-based feature selection are established. Finally, consideration is given to the important feature selection of the air quality data to the appropriate relevant features. For predicting similarities, the categorization is applied to the Air Quality Index (AQI), which estimates the significant classes like good, satisfactory, moderate, poor, very poor, and severe. To categorize air data into different types, a linear classifier is used. The Multiclass Support Vector IoT-based Air Pollution Forecast technique is used for the classification. Using categorized air data to forecast air pollution helps reduce error rates.

2.2 BTBSR-QWEBC Model

The Air Pollution prediction is improved by using the BTBSR-QWEBC model to get better accuracy and minimum time consumption. The proposed model includes preprocessing, Feature Selection and classification. Initially, the Bilateral discretized Z-wavelet transform is employed for preprocessing the input data which is used to eliminate the noise. Next, the Otsuka indexing Broken-stick regression method is used for relevant feature selection for minimizing the prediction of air pollution time. At last, the weighted emphasis boost method is used for the prediction of air pollution on the data classification using features. QWEBC is an ensemble classification method designed for changing weak learners into strong learners. The weakest learner's base classifier has a weak correlation with actual classification. The strong learner is an ensemble classifier designed for preventing accurate classification outcomes to collect the weak learner results. A kernelized support vector is employed by the weak learner. The ensemble method gathers the weak learners and gives the

outcomes of strong classification for reducing the quadratic error. With the help of this technique, Air Pollution prediction is improved using accuracy and it decreases time consumption.

BTBSR-QWEBC model is used for achieving the data preprocessing by converting raw data into readable format then the subsequent training scheme is used to enhance the performance. This model employs the bilateral discretized Z- wavelet transform which changes the format or values of data and changes it into clean and usable data. Next, the model is designed for attaining the feature selection with Otsuka indexing Broken-stick regression. The feature selection method is a significant phase in the construction of the machine learning method. The major process is designed for detecting the finest practicable set of features to construct the machine learning method. The result of relevant feature selection is designed for minimizing the prediction of air pollution time. Finally, the proposed model produces the classification for predicting air pollution with a Quadratic weighted emphasis boost technique is designed for improving accuracy and minimize the time involved in air pollution forecasting. The quadratic weighted emphasis boost method is the machine learning ensemble classification for converting weak learners into strong learners. A weak learner is a classifier that is problematic to give the true classification. The strong learner is a classifier that offers the true classification. It effectively classifies air quality data for predicting air pollution. It consumes more time for forecasting air pollution data with minimum and error rates. For better forecasting of air pollution with maximum accuracy and minimum time and error rate than model 2 and model 1, the next model is proposed.

2.3 DR-LSSV Model

The DR-LSSV is intended for the accurate prediction of air pollution via IoT networks. One input layer, three hidden layers, and one output layer are the different layers that make up the DR-LSSV model. The input layer first accepts data on air quality as input. The noise is then removed using the first hidden layer of the Discretized Hartley Transformation (DHT)-based preprocessing model to convert the real inputs into real outputs. When using the Constrained Maximum Likelihood Linear Regression-based Feature Selection approach, the second hidden layer on the correlative feature significant features is more pertinent. Finally, the Concordance Correlative Least Square Support Vector-based Classification model within the third hidden layer achieves air pollution prediction findings, and the output is supplied to the output layer.

Preprocessing is a significant issue with environmental data that has a big impact on data analysis methods and how well machine learning approaches predict air pollution. To remove the noise from the data and obtain the true trends and patterns, the recurring air quality data is detected. The Discretized Hartley Transformation-based Preprocessing strategy is employed with the goal of achieving the comparable results of the proposed method. The real inputs are sent to the real outputs using a proposed preprocessing approach. Concentrating on the contaminants most responsible for air pollution is crucial for the development of effective air pollution control strategies. During preprocessing, noise data from air quality data is eliminated and achieves preprocessed air data to forecast pollution.

The process of preprocessing using the proposed technique is explained in the provided Algorithm 1.

Input: Dataset 'DS', IoT Devices or Sensors '$S = S_1, S_2, \dots, S_n$', Features '$F = F_1, F_2, \dots, F_n$', Air Quality data '$D = D_1, D_2, \dots, D_n$'

Output: processed air quality data 'PD'

Step 1: **Initialize** rows and columns of the air quality data 'P' and 'Q'

Step 2: **Begin**

Step 3: **For** each Dataset 'DS' with Sensors 'S' and air quality data '$D = D_1, D_2, \dots, D_n$'

Step 4: Obtain vector matrix separately for corresponding station based on day 'Day' and hour '$Hour$'

Step 5: Formulate two-dimensional HT pair

Step 6: Obtain discretized HT pair to transform real inputs to real outputs

Step 7: **Return** processed air quality data 'PD'

Step 8: **End for**

Step 9: **End**

Algorithm 1: DHT based Preprocessing

To obtain preprocessed data, the transformation-based preprocessing is explained. For the decision-making process and forecasting of air quality issues, the selection of pertinent features is a key stage. In order to simplify and improve the effectiveness of the detection, the number of dimensions within processed air quality data are reduced using various feature selection strategies. Here, the second hidden layer's Constrained Maximum Likelihood Linear Regression (CMLLR) is used for the feature selection process.

The feature section procedure used to choose pertinent features from air data is described in Algorithm 2. The most widely used method for selecting features while considering the relevance of correlating features is constrained maximum likelihood linear regression. It is used to apply a higher likelihood function to explain the most pertinent or meaningful air quality index (AQI) forecast. Pollutant levels in the air can be preserved with the aid of AQI. To better anticipate air pollution, several categorization systems have been improved. However, AQI limitations due to their nature and inappropriate classification may have a negative impact on human health. In order to accurately anticipate air pollution using suitable classification and reduce error rates, the suggested approach is required.

Input: Dataset 'DS', IoT Devices or Sensors '$S = S_1, S_2, \dots, S_n$', Features '$F = F_1, F_2, \dots, F_n$', Air Quality data '$D = D_1, D_2, \dots, D_n$'

Output: Computationally efficient feature selection 'FS'

Step 1: **Initialize** processed air quality data 'PD'

Step 2: **Initialize** arbitrary air quality sample data '$(U_1, U_2, \dots, U_{n0})$' (i.e., processed station data athourly basis

Step 3: **Initialize** arbitrary air quality sample data '$(V_1, V_2, \dots, V_{n0})$' (i.e., processed station data at daily basis)

Step 4: **Begin**

Step 5: **For** each Dataset 'DS' with Sensors 'S' and processed air quality data 'PD'

Step 6: Evaluate logistic regression function for the homogenized air quality sample data

Step 7: Evaluate maximum semi-parametric likelihood estimator

Step 8: **Return** features selected 'FS'

Step 9: **End for**

Step 10: **End**

Algorithm 2: CMLLR based Feature Selection

In the third hidden layer, the Concordance Correlative Least Square Support Vector (CCLSSV) based Classification model is employed for calculating the non-linear association between input variables and output variables using minimal error. From the measurement of the non-linear relationship among the input variables and output variables, an error rate or evaluating the exposure of the target population for certain air pollutants detection is said to be precise. Besides, the utilization of the CC function, the non-linear relationship between the input variables and output variables are recognized for attaining the target population to certain air pollutants.

At Algorithm 3, Initially, number of preprocessed data and selected features of data from dataset is considered as input. For each selected feature, least support vector is formulated to classify data with minimum error. After that, CC value is estimated by applying Kendall's Rank Correlation function. Based on the coefficient value, data is classified into particular class for attaining accurate prediction. Based on classified data, air pollutions are forecasted with higher accuracy and minimum error.

Input: Dataset 'DS', IoT Devices or Sensors '$S = S_1, S_2, \ldots, S_n$', Features '$F = F_1, F_2, \ldots, F_n$', Air Quality data '$D = D_1, D_2, \ldots, D_n$'

Output:

Step 1: **Initialize** processed air quality data 'PD', features selected 'FS'

Step 2: **Begin**

Step 3: **For** each Dataset 'DS' with Sensors 'S', processed air quality data 'PD' and features selected 'FS'

Step 4: Formulate Least Square Support Vector-based Classification

Step 5: Evaluate minimum correlation with respect to error

Step 6: Measure Kendall's Rank Correlation Coefficient

Step 7: **If** '$\tau\ lies\ between -1\ and -0.5$'

Step 8: **Then** air quality is very poor

Step 9: **End if**

Step 10: **If** '$\tau\ lies\ between -5\ and\ 0$'

Step 11: **Then** air quality is poor

Step 12: **End if**

Step 13: **If** '$\tau\ lies\ equals\ 0$'

Step 14: **Then** air quality is good

Step 15: **End if**

Step 16: **If** '$\tau\ lies\ between\ 0\ and +0.5$'

Step 17: **Then** air quality is satisfactory

Step 18: **End if**

Step 19: **If** '$\tau\ lies\ between +0.5\ and\ 1$'

Step 20: **Then** air quality is moderate

Step 21: **End if**

Step 22: **Else**

Step 23: Air quality is very severe

Step 24: **End if**

Step 25: **End for**

Algorithm 3: CCLSSV based Classification

3. Results and Discussions

The developed models are performed for attaining enhanced air pollution forecasting with higher accuracy and minimum error. The performance of proposed models is compared with two different existing methods. The experimental results of the proposed models are established with different performance metrics like air pollution forecasting accuracy, air pollution forecasting time, and error rate. The observed result confirms that the proposed models produce higher prediction accuracy and lesser forecasting time when compared to the conventional techniques.

3.1 Measure of forecasting accuracy

The accurate air quality prediction is employed to be important for systematic pollution control. Air pollution forecasting accuracy is measured in percentage (%).

Table 1: Measure of Air pollution forecasting accuracy

Number of air quality data	Air pollution forecasting accuracy (%)				
	Existing Deep-AIR framework	Existing IMD-VAE	Proposed LR-MSV	Proposed BTBSR-QWEBC	Proposed DR-LSSV
20,000	82.25	76.25	83.62	85.25	87.96
40,000	83.66	77.52	84.25	86.62	88.76
60,000	83.93	77.62	84.68	87.63	89.96
80,000	84.22	78.56	85.63	88.25	91.01
1,00,000	85.63	79.52	86.11	89.64	91.22
1,20,000	86.89	80.04	87.65	90.25	91.52
1,40,000	87.61	80.25	88.62	90.63	92.04
1,60,000	88.11	80.63	89.52	91.25	92.66
1,80,000	88.24	81.72	89.82	91.36	92.81
2,00,000	89.62	82.46	90.36	91.67	93.66

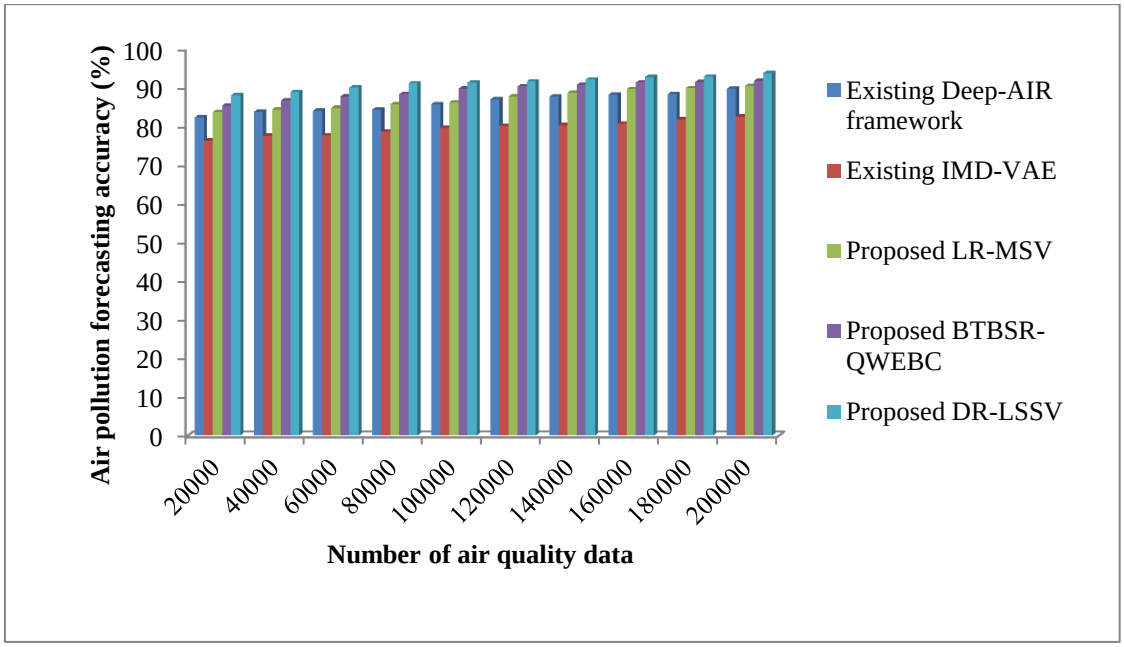

Figure 2: Graphical representation of Air pollution forecasting accuracy

Figure 2 demonstrates the air pollution forecasting accuracy obtained by the three proposed models, namely, LR-MSV, BTBSR-QWEBC, DR-LSSV and two existing techniques. Figure 2 gives the sample outcomes of air pollution accuracy for the air quality data in India. From that, the proposed DR-LSSV method attains higher accuracy on air pollution prediction and predicts the multiple classes on the Kendall's Rank Correlation Coefficient. Differences among any pair of observations within the output prediction and ground-truth forecasting to every different air quality data which varies between 20,000 to 2,00,000 data are considered. Then, the accurate air pollution prediction can be obtained and is considered to control the dangerous factor caused by air pollution, and therefore results in a healthy situation.

3.2 Measure of forecasting time

Secondly, it aims to classify the performance of the proposed models using the prediction of air pollution time used for forecasting air pollution. The air pollution forecasting time is calculated in milliseconds (ms).

Table 2: Measure of Air pollution forecasting time

Number of air quality data	Air pollution forecasting time (ms)				
	Existing Deep-AIR framework	Existing IMD-VAE	Proposed LR-MSV	Proposed BTBSR-QWEBC	Proposed DR-LSSV
20,000	2120	2500	2000	1680	1360
40,000	2060	2420	1960	1620	1300
60,000	2000	2350	1860	1550	1250
80,000	1940	2260	1820	1510	1200
1,00,000	1890	2150	1800	1480	1150
1,20,000	1800	2070	1770	1450	1100
1,40,000	1750	2020	1700	1420	1000
1,60,000	1700	1980	1680	1400	950
1,80,000	1680	1800	1630	1350	900
2,00,000	1650	1750	1600	1300	860

Figure 3 demonstrate the impact of air pollution forecasting time with various numbers of air quality data varying from 20,000 to 2,00,000. The simulation is achieved using the performance of various thresholds (i.e., stations being observed both daily and hourly basis). Trends of the prediction time involved on a daily and hourly basis can be considered. The amount of air quality data is improved, the percentage of maximum semi-parametric likelihood estimator is reduced, and the percentage of air pollution forecasting time decreased slowly. The reason behind this is the homogenized air quality sample of air quality data enhances using the growth of the improperly processed outcomes.

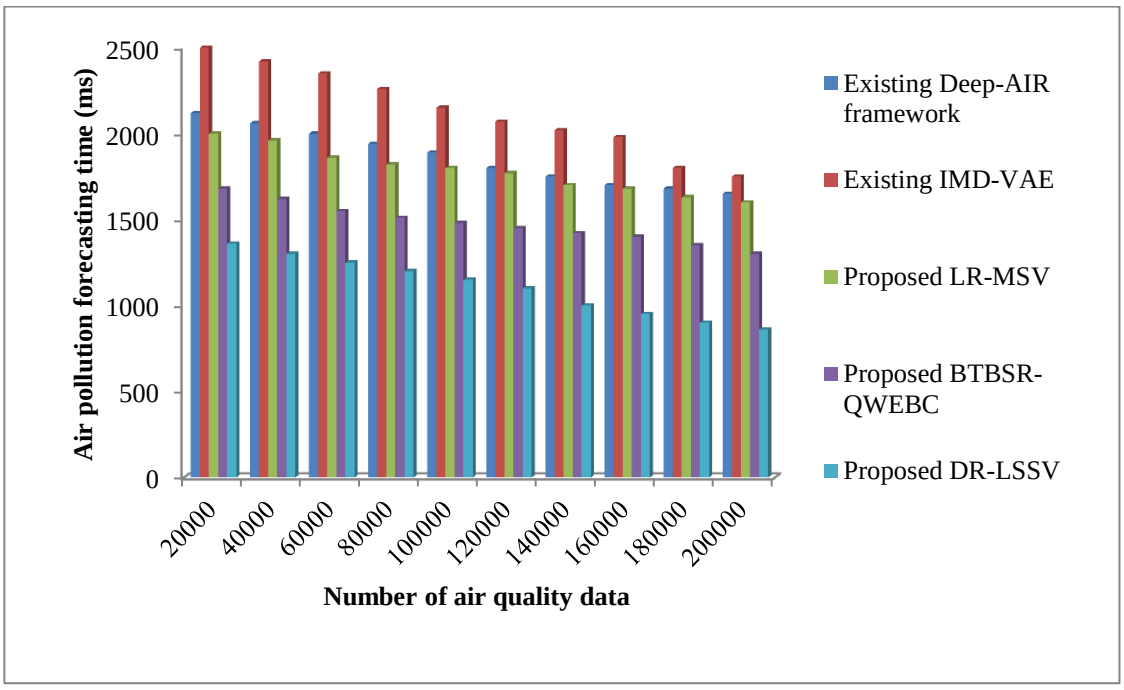

Figure 3: Graphical representation of Air pollution forecasting time

3.3 Measure of Error Rate

At last, the error rate must be calculated for air pollution forecasting process. The error rate is mathematically expressed in terms of percentage (%).

Table 3: Measure of Air pollution forecasting Error rate

Number of air quality data	Error Rate (%)				
	Existing Deep-AIR framework	Existing IMD-VAE	Proposed LR-MSV	Proposed BTBSR-QWEBC	Proposed DR-LSSV
20,000	17.75	23.75	16.38	14.75	12.04
40,000	16.34	22.48	15.75	13.38	11.24
60,000	16.07	22.38	15.32	12.37	10.04
80,000	15.78	21.44	14.37	11.75	8.99
1,00,000	14.37	20.48	13.89	10.36	8.78
1,20,000	13.11	19.96	12.35	9.75	8.48
1,40,000	12.39	19.75	11.38	9.37	7.96
1,60,000	11.89	19.37	10.48	8.75	7.34
1,80,000	11.76	18.28	10.18	8.64	7.19
2,00,000	10.38	17.54	9.64	8.33	6.34

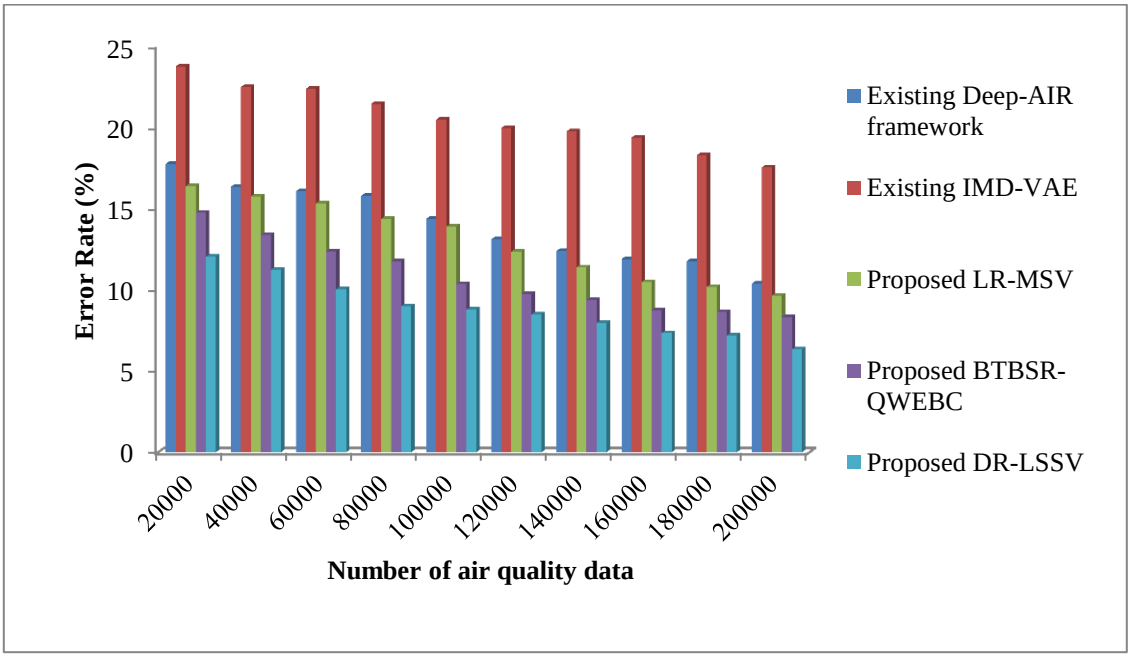

Figure 4 Graphical representation of Air pollution forecasting error rate

Figure 4 illustrates the error rate using the various air quality data attained by various time instances varies from 20,000 to 2,00,000. From the error rate, it enhanced with the improving air quality samples comparatively minimized using the proposed model upon comparison with existing techniques. The lesser error rate contribution using the DR-LSSV model was used for the utilization of the Multiclass Support Vector IoT-based Air Pollution Forecast.

4. Conclusion

The proposed models LR-MSV, BTBSR-QWEBC, and DR-LSSV provides enhanced prediction of the air quality data with higher forecasting accuracy and minimum time. Initially, the LR-MSV model is presented to forecast air pollution based on air quality index value. The wavelet sliding window process removed noise air data and attains preprocessed data. With obtained data, regression-based feature selection is carried out to select significant features of air data. Then, support vector classification is performed to classify data for forecasting air pollution with lesser time. Secondly, the BTBSR-QWEBC model is developed to achieve minimized time and memory consumption during air pollution forecasting. The noise air data is eliminated using a bilateral discretized process. Next, the Otsuka correlation coefficient value is estimated to select relevant features of data. With selected features, boosting technique is carried out for improving the accuracy of air pollution forecasting. Finally, the DR-LSSV model is explained for efficient forecasting of air pollution with a reduced error rate. The air data with noise removed and selected relevant features of data are considered for classifying data. By applying Kendall's Rank Correlation Coefficient, data is classified to forecast air pollution efficiently. Proposed models are provided to experiment results evaluated in different parameters such as error rate, air pollution forecasting time, and air pollution forecasting accuracy when compared to the other state of method.

References

[1] Qi Zhang, Yang Han, Victor O. K. Li, Jacqueline C. K. Lam, "*Deep-AIR: A Hybrid CNN-LSTM Framework for Fine-Grained Air Pollution Estimation and Forecast in Metropolitan Cities*", IEEE Access, Volume 10, 2022, 10.1109/ACCESS.2022.3174853.

[2] AbdelkaderDairi, FouziHarrou, SofianeKhadraoui, Ying, "*Integrated Multiple Directed Attention-based Deep Learning for Improved Air Pollution Forecasting*", IEEE Transactions on Instrumentation and Measurement, Volume 70, 2021, DOI: 10.1109/TIM.2021.3091511.

[3] K. Krishna Rani Samal, Ankit Kumar Panda, KorraSathyaBabu, Santos Kumar Das, "*An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach*", Sustainable Cities and Society, Elsevier, Volume 70, 2021, Pp. 1-13. https://doi.org/10.1016/j.scs.2021.102923.

[4] AzimHeydari, MeysamMajidiNezhad, DavideAstiaso Garcia, FarshidKeynia&Livio De Santoli, "*Air pollution forecasting application based on deep learning model and optimization algorithm*", Clean Technologies and Environmental Policy, Springer, Volume 24, 2022, Pp. 607–621. https://doi.org/10.1007/s10098-021-02080-5.

[5] Gao Huang, ChunjiangGe, TianyuXiong, Shiji Song, Le Yang, Baoxian Liu, Wenjun Yin, and Cheng Wu, "*Large scale air pollution prediction with deep convolutional networks*", Science China Information Sciences, Elsevier, Volume 64, 2021, Pp. 1-11. https://doi.org/10.1007/s11432-020-2951-1.

[6] AbdellatifBekker, BadrHasina, Samira Douzi& Khadija Douzi, "*Air-pollution prediction in smart city, deep learning approach*", *Journal of Big Data*, Springer, Volume 8, 2021, Pp. 1-21.

[7] Bingchun Liu, Xiaogang Yu, Jiali Chen, Qingshan Wang, "*Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model*", Atmospheric Pollution Research, Elsevier, Volume 12, 2021, Pp. 1-11. https://doi.org/10.1016/j.apr.2021.101144.

[8] Philipp Hähnel, JakubMareˇcek, JulienMonteil, FearghalO'Donncha, "*Using deep learning to extend the range of air pollution monitoring and forecasting*", *Journal of Computational Physics*, Elsevier, Volume 408, 2020, Pp. 1-13, https://doi.org/10.1016/j.jcp.2020.109278.

[9] Doreswamy, Harishkumar K S, Yogesh KM, Ibrahim Gad, "*Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models*", Procedia Computer Science, Elsevier, Volume 171, 2020, Pp. 2057–2066. https://doi.org/10.1016/j.procs.2020.04.221.

[10] Jianzhou Wang, Hongmin Li, Huang Yang, Ying Wang, "*Intelligent multivariable air-quality forecasting system based on feature selection and modified evolving interval type-2 quantum fuzzy neural network*", Environmental Pollution, Elsevier, Volume 274, 2021, Pp. 1-17. https://doi.org/10.1016/j.envpol.2021.116429.

[11] Shengdong Du; Tianrui Li; Yan Yang; Shi-Jinn Horng, "*Deep Air Quality Forecasting Using Hybrid Deep Learning Framework*", IEEE Transactions on Knowledge and Data Engineering, Volume 33, Issue 6, 2021, Pp. 2412 – 2424. DOI: 10.1109/TKDE.2019.2954510.

[12] Ning Jin; YongkangZeng; Ke Yan; ZhiweiJi, "*Multivariate Air Quality Forecasting with Nested Long Short-Term Memory Neural Network*", IEEE Transactions on Industrial Informatics, Volume 17, Issue 12, 2021, Pp.8514 – 8522. DOI: 10.1109/TII.2021.3065425.

[13] S Abirami and, P Chitra, "*Regional air quality forecasting using spatiotemporal deep learning*", *Journal of Cleaner Production*, Elsevier, Volume 283, 2021, Pp. 1-14. https://doi.org/10.1016/j.jclepro.2020.125341.

[14] DewenSeng, Qiyan Zhang, Xuefeng Zhang, Guangsen Chen and Xiyuan Chen, "*Spatiotemporal prediction of air quality based on LSTM neural network*", Knowledge-Based Systems, Elsevier, Volume 233, 2021, Pp. 1-12. https://doi.org/10.1016/j.aej.2020.12.009.

[15] Yu Huang, Josh Jia-Ching Ying, and Vincent S. Tseng, "*Spatio-attention embedded recurrent neural network for air quality prediction*", Knowledge-Based Systems, Elsevier, Volume 233, 2021, Pp. 1-14. https://doi.org/10.1016/j.knosys.2021.107416.

[16] XiangyuZou, JinjinZhao, Duan Zhao, Bin Sun,1,2 Yongxin He, and Stelios Fuentes, "*Air Quality Prediction Based on a Spatiotemporal Attention Mechanism*", Mobile Information Systems, Hindawi, Volume 2021, February 2021, Pp. 1-12. https://doi.org/10.1155/2021/6630944.

[17] Yuan Huang, Yuxing Xiang, Ruixiao Zhao, Zhe Cheng, "*Air Quality Prediction Using Improved PSO-BP Neural Network*", IEEE Access, Volume 8, 2020, Pp.99346 – 99353. DOI: 10.1109/ACCESS.2020.2998145.

[18] Hongqiao Chen, MengxiGuan, Hui Li, "*Air Quality Prediction Based on Integrated Dual LSTM Model*", IEEE Access, Volume 9, 2021, Pp. 93285 – 93297. DOI: 10.1109/ACCESS.2021.3093430.

[19] JianxianCai, Xun Dai, Li Hong, ZhitaoGao, and ZhongchaoQiu, "*An Air Quality Prediction Model Based on a Noise Reduction Self-Coding Deep Network*", Mathematical Problems in Engineering, Hindawi, Volume 2020, May 2020, Pp. 1-12.

[20] Shan Jiang, Zu-Guo Yu, Vo V. Anh, Yu Zhou, "*Long- and short-term time series forecasting of air quality by a multiscale framework*", Environmental Pollution, Elsevier, Volume 271,2021, Pp. 1-10. https://doi.org/10.1016/j.envpol.2020.116381.

PUBLICATIONS

International Journals:

1. M. Dhanalakshmi & Dr. V. Radha. (2023), "Novel Regression and Least Square Support Vector Machine Learning Technique for Air Pollution Forecasting", SSRG International Journal of Engineering Trends and Technology, Vol. 71, Issue. 04, pp: 147-158, ISSN: 2231-5381. (Scopus Indexed)
2. M. Dhanalakshmi & Dr. V. Radha. (2022), "Discretized Linear Regression and Multiclass Support Vector based Air pollution Forecasting Technique", SSRG International Journal of Engineering Trends and Technology, Vol. 70, Issue. 11, pp: 315-323, ISSN: 2231-5381. (Scopus Indexed)
3. M. Dhanalakshmi & Dr. V. Radha. (2021), "Vehicular Air Purifier – IoT Enabled System with Artificial Intelligence to Prevent Air Pollution", International Research Journal on Advance Science Hub, Vol. 03, Issue.07S, pp: 75-78, ISSN: 2582-4376. (peer-reviewed)
4. M. Dhanalakshmi & Dr. V. Radha (2022), "Air Pollution Forecasting using Novel Feature Selection and Classification Model", International Journal of Engineering Technology and Management Sciences, Vol. 06, Issue. 6, pp.: 20-27, ISSN: 2581-4621. (peer-reviewed)
5. M. Dhanalakshmi & Dr. V. Radha. (2021), "A Survey Paper on Vehicles Emitting Air Quality and Prevention of Air Pollution by using IoT along with Machine Learning

Approaches", Turkish Journal of Computer and Mathematics Education, Vol. 12, No. 11, pp: 5950-5962, ISSN: 1309-4653. (peer-reviewed)

Book Chapters and International Conference:

1. M. Dhanalakshmi & Dr. V. Radha, (2022)," Enhanced Preprocessing Technique for Air Pollution Forecasting System Using Big Data and Internet of Things ", Data Intelligence and Cognitive Informatics, ISBN 978-981-19-6003-1, DOI: 10.1007/978-981-19-6004-8. **(Springer – Scopus Indexed)**
2. M. Dhanalakshmi & Dr. V. Radha, (2022), "Comparative Study on IoT based System for Enhancing Air Pollution Forecasting Accuracy and Air Quality Index Monitoring", International Conference on Emerging Trends in IoT and Computing Technologies. ISBN 9781003350057. DOI: 10.1201/9781003350057 **(CRC Press Taylor and Francis Group –Scopus Indexed)**
3. M. Dhanalakshmi & Dr. V. Radha, (2022), "Novel Regression and Feature Selection Techniques for Air Pollution Forecast with IoT", Second International Conference on Emerging Trends on commerce, Science, Engineering, Management and Technology (ICCSEMT - II), pp:109. (**Conference Proceedings)**